■12 群（電子情報通信基礎）－ 3 編（統計・確率）

2 章　確率変数と確率分布関数

（執筆者：加藤　剛）[2009 年 3 月 受領]

■概要■

可測空間を構成する集合は一般には任意の集合でかまわないが，確率を考えるときは，数字からなる集合にしておいた方が扱いやすい．そこで，可測空間を定義域とし，実数または実ベクトルを対応させる関数もしくは写像を考える．前者を確率変数，後者を確率ベクトルという．確率変数及び確率ベクトルに対しては，もとになる可測空間上で定義された確率から確率分布が導かれ，更に確率分布からは，確率分布関数，確率関数，確率密度関数という概念が導き出される．また，3 編 1 章で定めた事象の独立性と同様にして，確率変数や確率ベクトルに対しても独立性が定められる．そして，独立性の性質は，確率分布関数，確率関数，確率密度関数によっても記述することができる．

確率変数は，確率分布で規定されるある種の規則に従って，複数回の観測を行うと不規則な値の系列を与えるが，それらの値の平均や値のばらつき具合は，期待値や分散といった指標で数値的にとらえられる．他方，期待値や分散の計算方法を活用すると，特性関数や積率母関数という概念を定めることができる．これらの関数は高次の積率の計算に使えるばかりでなく，確率分布の同一性や確率分布関数の収束を判定する道具としても利用可能である．

章の最後では，確率関数や確率密度関数を使った条件付き確率の定義を行い，条件付き期待値の概念を導く．

【本章の構成】

本章では，確率変数（2–1 節），確率分布関数（2–2 節），確率関数と確率密度関数（2–3 節），確率変数の独立性（2–4 節），期待値（2–5 節），特性関数と積率母関数（2–6 節），条件付き分布（2–7 節）の各概念の定義を述べ，関連する諸性質を紹介する．

■12 群 - 3 編 - 2 章

2-1 確率変数

（執筆者：加藤　剛）［2009 年 3 月 受領］

確率変数とは，直感的にいえば，観測を行うたびごとに不規則にとる値を変える変数のことである．例えば，X をバス停におけるバスの待ち時間を表す変数とするとき，X は調査をするたびごとに異なる値をとるであろうし，しかも，値の取り方は不規則に変化するであろう．したがって，X は確率変数とみなせる．そして，このような確率変数を n 個並べたベクトル $X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$ を確率ベクトルと呼ぶことも自然であろう．

確率変数と確率ベクトルの厳密な定義は，次のとおりである．実数全体の集合 R に対し，R のあらゆる区間全体からなる集合族を $\mathscr{A}$ とする．このとき，$\mathscr{A}$ から生成される σ-集合体を R 上の**ボレル集合体** (Borel field) といい，記号 $\mathscr{B}$ で表す．すなわち，$\mathscr{B} = \sigma(\mathscr{A})$ である．可測空間 $(\Omega, \mathscr{F})$ を前提として，Ω 上で定義された関数 $X : \Omega \to R$ が，すべての $B \in \mathscr{B}$ について

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathscr{F}$$

を満たすとき，$X = X(\omega)$ を $(\Omega, \mathscr{F})$ 上で定義された**確率変数** (random variable) と呼ぶ.

また，n 次元ユークリッド空間 R^n のあらゆる開集合からなる集合族を $\mathscr{A}$ とするとき，$\mathscr{A}$ から生成される σ-集合体を R^n 上の**ボレル集合体** (Borel field) といい，$\mathscr{B}^n$ で表す．可測空間 $(\Omega, \mathscr{F})$ を前提として，Ω 上で定義された写像 $X : \Omega \to R^n$ が，すべての $B \in \mathscr{B}^n$ について

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathscr{F}$$

を満たすとき，$X = X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$ を $(\Omega, \mathscr{F})$ 上で定義された n 次元**確率ベクトル** (random vector) と呼ぶ.

$(\Omega, \mathscr{F}, P)$ を確率空間，X を $(\Omega, \mathscr{F})$ 上で定義された確率変数とする．このとき，$B \in \mathscr{B}$ に対して

$$P^X(B) = P(X^{-1}(B))$$

と定めると，P^X は可測空間 $(R, \mathscr{B})$ 上の確率（測度）になる．この P^X を X の**確率分布** (probability distribution) と呼ぶ．n 次元確率ベクトル $X = (X_1, \ldots, Xn)$ についても，

$$P^X(B) = P(X^{-1}(B)), \quad B \in \mathscr{B}^n,$$

によって，可測空間 $(R^n, \mathscr{B}^n)$ 上で X の確率分布 P^X を定義することができる.

■12 群 - 3 編 - 2 章

2-2 確率分布関数

（執筆者：加藤　剛）［2009 年 3 月 受領］

前節で定めた確率分布をもとに，確率分布関数という概念を定める．

定義 2–2.1. $(\Omega,\mathscr{F},P)$ と $(R,\mathscr{B},P^X)$ を確率空間とする．ここで，P^X は $(\Omega,\mathscr{F})$ 上で定義された確率変数 X の確率分布である．このとき，

$$
\begin{aligned}
F_X(x) &= P\{\omega\in\Omega : X(\omega)\le x\} = P(X^{-1}((-\infty,x])), \quad (-\infty,x]\in\mathscr{B}, \\
&= P^X((-\infty,x]), \quad x\in R,
\end{aligned}
$$

によって定義される実数値関数 $F_X(x)$ を, X または P^X の**確率分布関数** (probability distribution function) と呼ぶ.

確率分布関数は，次の性質を持つ．

定理 2–2.1. $F_X(x)$ を，確率変数 X の確率分布関数とする．

(1) $0\le F_X(x)\le 1$.

(2) $x_1<x_2 \ \Rightarrow\ F_X(x_1)\le F_X(x_2)$.　（単調非減少）

(3) $\displaystyle\lim_{x\to\infty}F_X(x)=1, \quad \lim_{x\to-\infty}F_X(x)=0$.

(4) $\displaystyle\lim_{h\to+0}F_X(x+h)=F_X(x)$.　（右連続）

証明 (1) 確率分布関数の定義と確率の性質から明らか.

(2) $x_1<x_2 \ \Rightarrow\ F_X(x_1)=P^X((-\infty,x_1])\le P^X((-\infty,x_2])=F_X(x_2)$.

(3) $A_n=\{\omega : X(\omega)\le n\}$ とおくと，$\{A_n\}_{n=1}^{\infty}$ は単調増加な集合列で，$\displaystyle\bigcup_{n=1}^{\infty}A_n=\Omega$．そこで，3 編 1 章定理**??**の (2) より，

$$
\lim_{x\to\infty}F_X(x)=\lim_{n\to\infty}F_X(n)=\lim_{n\to\infty}P(A_n)=P\left(\bigcup_{n=1}^{\infty}A_n\right)=P(\Omega)=1.
$$

また，$B_n=\{\omega : X(\omega)\le -n\}$ とおくと，B_n は単調減少で，$\displaystyle\bigcap_{n=1}^{\infty}B_n=\emptyset$ である．よって，3 編 1 章定理**??**の (3) より

$$
\lim_{x\to-\infty}F_X(x)=\lim_{n\to\infty}F_X(-n)=\lim_{n\to\infty}P(B_n)=P\left(\bigcap_{n=1}^{\infty}B_n\right)=P(\emptyset)=0.
$$

(4) $x\in R$ を任意にとって固定し，$A_n=\left\{\omega : X(\omega)\le x+\dfrac{1}{n}\right\}$ と定める．すると，$\{A_n\}_{n=1}^{\infty}$ は

単調減少で，$\bigcap_{n=1}^{\infty} A_n = \{\omega : X(\omega) \le x\}$ である．よって，3 編 1 章定理**??**の (3) より，

$$\lim_{h \to +0} F_X(x+h) = \lim_{n\to\infty} F_X\left(x + \frac{1}{n}\right) = \lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = P\{\omega : X(\omega) \le x\} = F_X(x). \qquad \square$$

確率変数の場合と同様にして，確率ベクトルに対しても確率分布関数を定めることができる．

定義 2–2.2. $(\Omega, \mathscr{F}, P)$ と $(R, \mathscr{B}, P^X)$ を確率空間とする．ただし，P^X は確率ベクトル $X = (X_1, \ldots, X_n)$ の確率分布である．このとき，任意の $x = (x_1, \ldots, x_n) \in R^n$ に対して

$$F_X(x) = P\{\omega \in \Omega : X_1(\omega) \le x_1, \ldots, X_n(\omega) \le x_n\} = P^X\left(\prod_{i=1}^{n} (-\infty, x_i]\right)$$

によって定義される実数値関数 $F_X(x)$ を，X または P^X の**同時確率分布関数** (joint probability distribution function) と呼ぶ.

確率ベクトルの確率分布関数に対しても，本章定理 2–2.1 と同様な次の結果が成り立つ．ただし，証明を略する（[4]参照）.

定理 2–2.2.

(1) $0 \le F_X(x) \le 1$.

(2) 任意の $I = \prod_{i=1}^{n} (a_i, b_i]$ に対して，$\Delta_I^n F_X(x) \ge 0$. ただし，$\Delta_I^n F_X(x)$ は $F_X(x)$ の区間 I 上における n 次差分

$$\begin{aligned} \Delta_I^n F_X(x) \quad &= \quad F_X(b_1, \ldots, b_n) - \{F_X(a_1, b_2, \ldots, b_n) + \cdots F_X(b_1, \ldots, b_{n-1}, a_n)\} \\ &\quad\ + \{F_X(a_1, a_2, b_3, \ldots, b_n) + \cdots + F_X(b_1, \ldots, b_{n-2}, a_{n-1}, a_n)\} \\ &\quad\ - \cdots + (-1)^n F_X(a_1, \ldots, a_n) \end{aligned}$$

を表す.

(3) $\displaystyle\lim_{\substack{x_i \to \infty \\ i=1,\ldots,n}} F_X(x_1, \ldots, x_n) = 1, \quad \lim_{x_i \to -\infty} F_X(x_1, \ldots, x_i, \ldots, x_n) = 0.$

(4) 任意の $(x_1, \ldots, x_n) \in R^n$ と $(\varepsilon_1, \ldots, \varepsilon_n) \in \prod_{i=1}^{n} \{0, 1\}$ に対し，

$$\lim_{\substack{h_i \to +0 \\ i=1,\ldots,n}} F_X(x_1 + \varepsilon_1 h_1, \ldots, x_n + \varepsilon_n h_n) = F_X(x_1, \ldots, x_n).$$

■12 群 - 3 編 - 2 章

2-3 確率関数と確率密度関数

（執筆者：加藤　剛）[2009 年 3 月 受領]

確率変数 X を一定時間内にセンサに衝突する電子の数とすると，X は離散的な値をとる．他方，X をバス停におけるバスの待ち時間とすると，X は連続的な値をとる．このように，確率変数には，大きく分けて 2 種類あることが分かる．

定義 2–3.1.

(i) $D=\{a_0,a_1,\ldots\}$ を有限集合または可算無限集合とし，$X(\omega)\in D$ であるとき，X を**離散型確率変数** (discrete type random variable) といい，X の確率分布 P^X を**離散型確率分布** (discrete type probability distribution) という．

(ii) D 上で定義される関数

$$p(a_k)=P\{\omega\in\Omega : X(\omega)=a_k\}=P^X(\{a_k\}),\quad k=0,1,\ldots,$$

を，X または P^X に対応する**確率関数** (probability function) と呼ぶ.

定義 2–3.2.

(i) 確率変数 X が R 上の連続的な値をとるとき, X を**連続型確率変数** (continuous type random variable) といい，X の確率分布 P^X を**連続型確率分布** (continuous type probability distribution) という.

(ii) P^X が連続型確率分布であるとき，任意の $a,b\in R$, $a<b$, に対して

$$P\{\omega\in\Omega : a<X(\omega)\le b\}=P^X((a,b])=\int_a^b f(x)\,dx$$

を満たす可積分関数 $f(x)$ を，X または P^X に対応する**確率密度関数** (probability density function) と呼ぶ.

本章定義 2–3.1 より，確率関数は次の性質を持つことが直ちに示せる．

(1) $p(a_k)\ge 0,\quad k=0,1,\ldots.$

(2) $\displaystyle\sum_{k=0}^{\infty} p(a_k)=1.$

(3) $\displaystyle F_X(x)=\sum_{\{k:a_k\le x\}} p(a_k).$

また，本章定義 2–3.2 より，確率密度関数は次の性質を持つことが分かる．

(1) 任意の $x\in R$ について，$f(x)\ge 0.$

(2) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1.$

(3) $f(x)$ の連続点 x において，$\dfrac{d}{dx}F_X(x) = f(x)$.

確率変数についての議論と同様にして，確率ベクトルについても確率関数と確率密度関数を定めることができる．

定義 2−3.3.

(i) $D = \{a_k = (a_{k1}, a_{k2}, \ldots, a_{kn}) : k = 0,1,\ldots\}$ を有限集合または可算無限集合とし，$X(\omega) = (X_1(\omega),\ldots,X_n(\omega)) \in D$ であるとき，X を n 次元**離散型確率ベクトル** (discrete type random vector) といい，X の確率分布 P^X を**離散型確率分布** (discrete type probability distribution) という．

(ii) D 上で定義される関数

$$p(a_k) = P\{\omega \in \Omega : X(\omega) = a_k\} = P^X(\{a_k\}), \quad k = 0,1,\ldots,$$

を，X または P^X に対応する**同時確率関数** (joint probability function) と呼ぶ．

定義 2−3.4.

(i) 確率ベクトル $X = (X_1,\ldots,X_n)$ の各要素 X_i が R 上の連続的な値をとるとき，X を n 次元**連続型確率ベクトル** (continuous type random vector) といい，X の確率分布 P^X を**連続型確率分布** (continuous type probability distribution) という．

(ii) P^X が連続型確率分布であるとき，任意の $A = \displaystyle\prod_{i=1}^{n}(a_i, b_i] \in \mathscr{B}^n$ に対して

$$P\{\omega \in \Omega : X(\omega) \in A\} = P^X(A) = \int \cdots \int_A f(x_1,\ldots,x_n)\,dx_1 \cdots dx_n$$

を満たす可積分関数 $f(x) = f(x_1,\ldots,x_n)$ を，X または P^X に対応する**同時確率密度関数** (joint probability density function) と呼ぶ．

2 次元の同時確率関数や同時確率密度関数と，1 次元の確率関数や確率密度関数の関係は，次の形で述べられる．

定義 2−3.5. $X = (X_1, X_2)$ を 2 次元確率ベクトルとする．

(i) $p_{X_1,X_2}(x_1,x_2)$ を X の同時確率関数とし，$X \in \{(a_i, b_j) : i, j = 0,1,\ldots\}$ であるとする．このとき，

$$\begin{aligned} p_{X_1}(x_1) &= \sum_{j=0}^{\infty} p_{X_1,X_2}(x_1, b_j), \\ p_{X_2}(x_2) &= \sum_{i=0}^{\infty} p_{X_1,X_2}(a_i, x_2) \end{aligned}$$

によって定められる $p_{X_1}(x_1)$ と $p_{X_2}(x_2)$ を, **周辺確率関数** (marginal probability function) と呼ぶ.

(ii) $f_{X_1,X_2}(x_1,x_2)$ を X の同時確率密度関数とするとき,

$$
\begin{aligned}
f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\,dx_2, \\
f_{X_2}(x_2) &= \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\,dx_1
\end{aligned}
$$

によって定められる $f_{X_1}(x_1)$ と $f_{X_2}(x_2)$ を, **周辺確率密度関数** (marginal probability density function) と呼ぶ.

ここでは確率ベクトルが 2 次元の場合を述べたが, 一般の n 次元確率ベクトルに対しても, 同様にして周辺確率関数や周辺確率密度関数が定義できる.

■12 群 - 3 編 - 2 章

2-4 確率変数の独立性

（執筆者：加藤　剛）［2009 年 3 月 受領］

この節では，確率変数の独立性と，それに関連する概念を紹介する．

定義 2–4.1.

(i) $(\Omega,\mathscr{F})$ と $(R,\mathscr{B})$ を可測空間とする．任意の $A,B\in\mathscr{B}$ に対し，$(\Omega,\mathscr{F})$ 上で定義された確率変数 X,Y が

$$P\{\omega : X(\omega)\in A, Y(\omega)\in B\} = P\{\omega : X(\omega)\in A\}P\{\omega : Y(\omega)\in B\}$$

を満たすとき，X と Y は**互いに独立** (mutually independent) であるという．

(ii) $(\Omega,\mathscr{F})$，$(R^n,\mathscr{B}^n)$，$(R^m,\mathscr{B}^m)$ を可測空間とする．任意の $A\in\mathscr{B}^n$，$B\in\mathscr{B}^m$ に対し，$(\Omega,\mathscr{F})$ 上で定義された n 次元確率ベクトル X と m 次元確率ベクトル Y が

$$P\{\omega : X(\omega)\in A, Y(\omega)\in B\} = P\{\omega : X(\omega)\in A\}P\{\omega : Y(\omega)\in B\}$$

を満たすとき，X と Y は**互いに独立** (mutually independent) であるという．

写像 $g : R^n \to R^p$ が，任意の $A\in\mathscr{B}^p$ に対して $g^{-1}(A) = \{x\in R^n : g(x)\in A\}\in\mathscr{B}^n$ を満たすとき，g を R^p 値ボレル可測関数と呼ぶ．いま，X を n 次元確率ベクトル，Y を m 次元確率ベクトル，g を R^p 値ボレル可測関数，h を R^q 値ボレル可測関数とする．このとき，

$$X \text{ と } Y \text{ が互いに独立} \Rightarrow g(X) \text{ と } h(Y) \text{ が互いに独立}$$

が成り立つ．このことは，次のようにして示される．

$$\begin{aligned}
P\{\omega : g(X(\omega))\in A,\ h(Y(\omega))\in B\} &= P\{\omega : X(\omega)\in g^{-1}(A),\ Y(\omega)\in h^{-1}(B)\}\\
&= P\{\omega : X(\omega)\in g^{-1}(A)\}P\{\omega : Y(\omega)\in h^{-1}(B)\}\\
&= P\{\omega : g(X(\omega))\in A\}P\{\omega : h(Y(\omega))\in B\}.
\end{aligned}$$

確率変数 X,Y の独立性とは，一口で言えば，X と Y の組 (X,Y) に関する事象の確率が，X のみの事象と Y のみの事象の確率の積として表されるということである．この考え方は，次の定理で示すように，確率分布関数，確率関数，確率密度関数に対してもあてはまる．確率変数の場合のみを記すが，確率ベクトルに対しても同様な結果が成り立つ（[4]参照）．

定理 2–4.1.

(1) $F_{X,Y}(x,y)$ を (X,Y) の同時確率分布関数，$F_X(x)$ を X の確率分布関数，$F_Y(y)$ を Y の確率分布関数とすると，

$$X \text{ と } Y \text{ が互いに独立} \Leftrightarrow \text{任意の } x,y\in R \text{ について，} F_{X,Y}(x,y) = F_X(x)F_Y(y).$$

(2) X と Y が離散型確率変数であり，$p_{X,Y}(x,y)$ を (X,Y) の同時確率関数，$p_X(x)$ を X の確率関数，$p_Y(y)$ を Y の確率関数，更に $(X,Y)\in D=\{(a_i,b_j):i,j=0,1,\ldots\}$ とすると，

$$X \text{ と } Y \text{ が互いに独立} \Leftrightarrow \text{任意の } (a_i,b_j)\in D \text{ について，} p_{X,Y}(a_i,b_j)=p_X(a_i)p_Y(b_j).$$

(3) X と Y が連続型確率変数であり，$f_{X,Y}(x,y)$ を (X,Y) の同時確率密度関数，$f_X(x)$ を X の確率密度関数，$f_Y(y)$ を Y の確率密度関数 とすると，

$$X \text{ と } Y \text{ が互いに独立} \Leftrightarrow \text{任意の } x,y\in R \text{ について，} f_{X,Y}(x,y)=f_X(x)f_Y(y).$$

証明 (1) X と Y が互いに独立ならば，任意の $x,y\in R$ について

$$\begin{aligned}F_{X,Y}(x,y) &= P\{X\in(-\infty,x],\ Y\in(-\infty,y]\}\\ &= P\{X\in(-\infty,x]\}P\{Y\in(-\infty,y]\}=F_X(x)F_Y(y).\end{aligned}$$

逆に，任意の $x,y\in R$ について $F_{X,Y}(x,y)=F_X(x)F_Y(y)$ が成り立つならば，任意の $(a_1,a_2]\times(b_1,b_2]\in\mathscr{B}^2$ について

$$\begin{aligned}P\{X\in(a_1,a_2],\ Y\in(b_1,b_2]\} &= P\{(X,Y)\in(a_1,a_2]\times(b_1,b_2]\}\\ &= F_{X,Y}(a_2,b_2)-F_{X,Y}(a_2,b_1)-F_{X,Y}(a_1,b_2)+F_{X,Y}(a_1,b_1)\\ &= F_X(a_2)F_Y(b_2)-F_X(a_2)F_Y(b_1)-F_X(a_1)F_Y(b_2)\\ &\quad +F_X(a_1)F_Y(b_1)\\ &= \{F_X(a_2)-F_X(a_1)\}\{F_Y(b_2)-F_Y(b_1)\}\\ &= P\{X\in(a_1,a_2]\}P\{Y\in(b_1,b_2]\}.\end{aligned}$$

このように，$(a_1,a_2]\times(b_1,b_2]$ という形をした任意の集合について結論が成り立てば，すべての $A\times B\in\mathscr{B}^2$ について $P\{X\in A,\ Y\in B\}=P\{X\in A\}P\{Y\in B\}$ が成り立つことも導ける．しかし，議論が煩雑になるので，その証明は省略する．
(2) X と Y が互いに独立ならば，任意の $(a_i,b_j)\in D$ について，

$$p_{X,Y}(a_i,b_j)=P\{X=a_i,Y=b_j\}=P\{X=a_i\}P\{Y=b_j\}=p_X(a_i)p_Y(b_j).$$

逆に，任意の $(a_i,b_j)\in D$ に対して $p_{X,Y}(a_i,b_j)=p_X(a_i)p_Y(b_j)$ が成り立つならば，任意の $A,B\in\mathscr{B}$ について

$$\begin{aligned}P\{X\in A,Y\in B\} &= \sum_{\{(i,j):(a_i,b_j)\in A\times B\}} p_{X,Y}(a_i,b_j)=\sum_{\{(i,j):(a_i,b_j)\in A\times B\}} p_X(a_i)p_Y(b_j)\\ &= \sum_{\{i:a_i\in A\}} p_X(a_i)\sum_{\{j:b_j\in B\}} p_Y(b_j)=P\{X\in A\}P\{Y\in B\}.\end{aligned}$$

(3) X と Y が互いに独立ならば，任意の $E=(a_1,a_2]\times(b_1,b_2]\in\mathscr{B}^2$ について，

$$\begin{aligned}\iint_E f_{X,Y}(x,y)\,dxdy &= P\{X \in (a_1,a_2],\, Y \in (b_1,b_2]\} = P\{X \in (a_1,a_2]\}P\{Y \in (b_1,b_2]\} \\ &= \int_{a_1}^{a_2} f_X(x)\,dx \int_{b_1}^{b_2} f_Y(y)\,dy = \iint_E f_X(x)\,f_Y(y)\,dxdy.\end{aligned}$$

したがって，被積分関数が等しいことになるので，$f_{X,Y}(x,y) = f_X(x)f_Y(y)$．逆に，任意の $x,y \in R$ について，$f_{X,Y}(x,y) = f_X(x)f_Y(y)$ ならば，任意の $(a_1,a_2] \times (b_1,b_2] \in \mathscr{B}^2$ に対して，

$$\begin{aligned}P\{X \in (a_1,a_2],\, Y \in (b_1,b_2]\} &= \int_{a_1}^{a_2}\int_{b_1}^{b_2} f_{X,Y}(x,y)\,dxdy = \int_{a_1}^{a_2}\int_{b_1}^{b_2} f_X(x)\,f_Y(y)\,dxdy \\ &= \int_{a_1}^{a_2} f_X(x)\,dx \int_{b_1}^{b_2} f_Y(y)\,dy = P\{X \in (a_1,a_2]\}P\{Y \in (b_1,b_2]\}.\end{aligned}$$

よって，X と Y は互いに独立になる． □

■12 群 - 3 編 - 2 章

2-5 期待値

（執筆者：加藤　剛）［2009 年 3 月 受領］

確率変数は，観測や試行を行うたびごとに，不規則に異なる値をとる．しかし，もしも観測や試行を繰り返したならば，確率変数のとる平均的な値というものが得られるであろう．この値のことを期待値あるいは平均と呼ぶ．この節では，期待値に関する定義と性質を述べる．なお，ここでは確率変数に関する結果しか紹介しないが，確率ベクトルに関しても期待値は定義され，以下で紹介するものと同様な性質を持つことを示せる（4) 参照）．

定義 2–5.1. X を可測空間 $(\Omega, \mathscr{F})$ 上の確率変数とし，X が離散型ならば $p(x)$, $x \in \{x_0, x_1, \ldots\}$, をその確率関数，$X$ が連続型ならば $f(x)$ をその確率密度関数とする．このとき，

$$E(X) = \begin{cases} \displaystyle\sum_{k=0}^{\infty} x_k p(x_k) & : \quad X \text{ が離散型のとき} \\[2ex] \displaystyle\int_{-\infty}^{\infty} x f(x)\, dx & : \quad X \text{ が連続型のとき} \end{cases}$$

で定義される $E(X)$ を X の**期待値** (expectation) または**平均** (mean) と呼ぶ.

確率変数の関数の期待値は，次の計算規則で計算できる．

定理 2–5.1. $g(x)$ が R 値ボレル可測関数であるとき，

$$E(g(X)) = \begin{cases} \displaystyle\sum_{k=0}^{\infty} g(x_k) p(x_k) & : \quad X \text{ が離散型のとき} \\[2ex] \displaystyle\int_{-\infty}^{\infty} g(x) f(x)\, dx & : \quad X \text{ が連続型のとき} \end{cases}$$

証明 X が離散型の場合を示す．$Y = g(X)$ とし，$X \in \{x_0, x_1, \ldots\}$ に対して Y がとる値を $\{y_0, y_1, \ldots\}$ とすると，Y の確率関数 $p_Y(y)$ は

$$p_Y(y_k) = \sum_{\{i: g(x_i) = y_k\}} p(x_i)$$

で与えられる．そこで，

$$E(Y) = \sum_{k=0}^{\infty} y_k p_Y(y_k) = \sum_{k=0}^{\infty} y_k \left(\sum_{\{i: g(x_i) = y_k\}} p(x_i) \right) = \sum_{i=0}^{\infty} g(x_i) p(x_i). \quad \square$$

期待値は，次の性質を持つ．

定理 2–5.2.

(1) 任意の $a \in R$ に対して，$E(X + a) = E(X) + a$.

(2) 任意の $a,b \in R$ に対して，$E(aX+bY)=aE(X)+bE(Y)$.

(3) X と Y が互いに独立 $\Rightarrow$ $E(XY)=E(X)E(Y)$.

(4) 任意の $\omega \in \Omega$ について $X(\omega) \geq 0$ $\Rightarrow$ $E(X) \geq 0$.

(5) 任意の $\omega \in \Omega$ について $X(\omega) \leq Y(\omega)$ $\Rightarrow$ $E(X) \leq E(Y)$.

証明 X,Y が連続型の場合を示す.

(1) $f_X(x)$ を X の確率密度関数とすると，

$$\begin{aligned} E(X+a) &= \int_{-\infty}^{\infty}(x+a)f_X(x)\,dx = \int_{-\infty}^{\infty} x f_X(x)\,dx + a\int_{-\infty}^{\infty} f_X(x)\,dx \\ &= E(X)+aP\{\omega : -\infty < X(\omega) < \infty\} = E(X)+a. \end{aligned}$$

(2) $f_{X,Y}(x,y)$ を X と Y の同時確率密度関数とすると，本章節で見たように，

$$f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy, \quad f_Y(y)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

は，それぞれ X と Y の周辺確率密度関数となる．この結果を利用すると，

$$\begin{aligned} E(aX+bY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by)\,f_{X,Y}(x,y)\,dxdy \\ &= a\int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\right)dx + b\int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\right)dy \\ &= a\int_{-\infty}^{\infty} x f_X(x)\,dx + b\int_{-\infty}^{\infty} y f_Y(y)\,dy \\ &= aE(X)+bE(Y). \end{aligned}$$

(3) $f_{X,Y}(x,y)$ を (X,Y) の同時確率密度関数，$f_X(x)$ と $f_Y(y)$ をそれぞれ X と Y の確率密度関数とする．本章定理 2–4.1 の (3) より，X と Y が互いに独立ならば，

$$f_{X,Y}(x,y)=f_X(x)\,f_Y(y).$$

よって，

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,f_{X,Y}(x,y)\,dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,f_X(x)\,f_Y(y)\,dxdy \\ &= \int_{-\infty}^{\infty} x f_X(x)\,dx \int_{-\infty}^{\infty} y f_Y(y)\,dy = E(X)E(Y). \end{aligned}$$

(4) 任意の $\omega \in \Omega$ について $X(\omega) \geq 0$ ならば，

$$E(X)=\int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{0}^{\infty} x f_X(x)\,dx \geq 0.$$

(5) $Z(\omega)=Y(\omega)-X(\omega)$ とおけば，任意の $\omega \in \Omega$ に対して $Z(\omega) \geq 0$. よって，(2) と (4) の結果より

$$0 \le E(Z) = E(Y - X) = E(Y) - E(X).$$

よって，$E(X) \le E(Y)$. □

本章定理 2–5.1 において特に $g(x) = x^k$（k は自然数）としたときの期待値 $E(X^k)$ を，k 次の**積率**（**モーメント**）(moment) と呼ぶ．1 次の積率が期待値になる．また，1 次と 2 次の積率を使って求められる量

$$\begin{aligned} V(X) &= E\{X - E(X)\}^2 \\ &= E(X^2) - \{E(X)\}^2 \end{aligned}$$

を X の**分散** (variance) と呼ぶ．直感的な解釈としては，期待値は確率分布全体の代表となるものである．また，分散は，確率変数 X について観測をしたとき，観測値のばらつき具合を表す指標となる．分散が小さいと，X の観測値は期待値の周りに集中して現れる．他方，分散が大きいと，X の観測値はまとまりがなく，ばらついた値をとる．

次に，2 つの確率変数の関連性を見る指標を定義する．

定義 2–5.2.

(i) 2 つの確率変数 X, Y に対して，

$$\begin{aligned} \mathrm{Cov}(X,Y) &= E[\{X - E(X)\}\{Y - E(Y)\}] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

で定められる量を，X と Y の**共分散** (covariance) という．

(ii) $V(X), V(Y)$ をそれぞれ X, Y の分散，$\mathrm{Cov}(X,Y)$ を X と Y の共分散とするとき，

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)V(Y)}}$$

を X と Y の**相関係数** (correlation coefficient) と呼ぶ．

分散，共分散，相関係数について，次の性質が成り立つ．

定理 2–5.3.

(1) $a \in R$ を定数とするとき，$V(X+a) = V(X)$.

(2) X と Y が互いに独立 $\Rightarrow$ $\mathrm{Cov}(X,Y) = 0$.

(3) $\{a_i\}$ を実数列とするとき，

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V(X_i) + \sum_{\{(i,j):i \neq j\}} a_i a_j \mathrm{Cov}(X_i, X_j).$$

特に $X_1,\ldots,X_n$ が互いに独立ならば,

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V(X_i).$$

(4) $|\rho(X,Y)| \le 1.$

証明 (1) 本章定理 2–5.2 の (1) より $E(X+a) = E(X)+a$ であるから,

$$V(X+a) = E\{(X+a)-E(X+a)\}^2 = E\{X-E(X)\}^2 = V(X).$$

(2) 本章定理 2–5.2 の (3) より，X とが独立ならば $E(XY) = E(X)E(Y)$ であるから,

$$\mathrm{Cov}(X,Y) = E(XY)-E(X)E(Y) = E(X)E(Y)-E(X)E(Y) = 0.$$

(3) 本章定理 2–5.2 の (2) より $E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$ なので,

$$\begin{aligned}
V\left(\sum_{i=1}^{n} a_i X_i\right) &= E\left\{\sum_{i=1}^{n} a_i X_i - E\left(\sum_{i=1}^{n} a_i X_i\right)\right\}^2 \\
&= E\left\{\sum_{i=1}^{n} a_i (X_i - E(X_i))\right\}^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E[\{(X_i - E(X_i)\}\{X_j - E(X_j\}] \\
&= \sum_{i=1}^{n} a_i^2 E\{X_i - E(X_i)\}^2 + \sum_{\{(i,j):i\neq j\}} a_i a_j E[\{X_i - E(X_i)\}\{X_j - E(X_j)\}] \\
&= \sum_{i=1}^{n} a_i^2 V(X_i) + \sum_{\{(i,j):i\neq j\}} a_i a_j \mathrm{Cov}(X_i, X_j).
\end{aligned}$$

特に $X_1,\ldots,X_n$ が互いに独立ならば，(2) より $\mathrm{Cov}(X_i,X_j) = 0$ $(i \neq j)$ なので,

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V(X_i).$$

(4) t を実変数として $U(\omega) = [t\{X(\omega)-E(X)\} - \{Y(\omega)-E(Y)\}]^2$ とおくと，任意の $\omega \in \Omega$ に対して $U(\omega) \ge 0$. そこで，本章定理 2–5.2 の (4) を適用して,

$$\begin{aligned}
0 \le E(U) &= t^2 E\{X-E(X)\}^2 - 2tE[\{X-E(X)\}\{Y-E(Y)\}] + E\{Y-E(Y)\}^2 \\
&= t^2 V(X) - 2t\,\mathrm{Cov}(X,Y) + V(Y).
\end{aligned}$$

これを t に関する 2 次不等式とみて判別式の符号を考えると,

$$\{\mathrm{Cov}(X,Y)\}^2 - V(X)V(Y) \le 0.$$

よって，$|\mathrm{Cov}(X,Y)| \leq \sqrt{V(X)V(Y)}$ となるので，$|\rho(X,Y)| \leq 1$ が得られる．　$\square$

本章定理 2–5.3 の (4) において，等号が成立するのは $Y = aX + b$ $(a \neq 0, b \in R)$ という関係が成り立つときである．つまり，相関係数とは，2 つの確率変数 X,Y について 1 次式の関係がどの程度成り立っているかを判定する指標となっている．X と Y が完全な 1 次式の関係にあるならば，$\rho(X,Y) = \pm 1$ となる．

■12 群 - 3 編 - 2 章

2-6 特性関数と積率母関数

（執筆者：加藤　剛）[2009 年 3 月 受領]

本章節において，k 次の積率を取り上げた．k 次の積率は本章定理 2–5.1 に基づいて計算してもよいが，次に紹介する特性関数を用いても求めることができる．

定義 2–6.1.　X を確率変数とするとき，

$$\varphi_X(t) = E(e^{itX}) = \begin{cases} \displaystyle\sum_{i=0}^{\infty} e^{itx_k} p(x_k) & : \quad X \text{ が離散型のとき} \\ \displaystyle\int_{-\infty}^{\infty} e^{itx} f_X(x)\,dx & : \quad X \text{ が連続型のとき} \end{cases}$$

を X の**特性関数** (characteristic function) と呼ぶ．ここで，$i = \sqrt{-1}$ である．また，X が離散型の場合，$X \in \{x_0, x_1, \ldots\}$ としている．

特性関数は，次の性質を持つ．

定理 2–6.1.　$\varphi_X(t), \varphi_Y(t)$ をそれぞれ X, Y の特性関数とする．このとき，$X+Y$ の特性関数 $\varphi_{X+Y}(t)$ について，

$$X \text{ と } Y \text{ が互いに独立} \ \Rightarrow\ \varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t).$$

証明　本章節で述べたことから，X と Y が互いに独立ならば，それぞれの関数である e^{itX} と e^{itY} も互いに独立になる．したがって，

$$\varphi_{X+Y}(t) = E(e^{it(X+Y)}) = E(e^{itX}e^{itY}) = E(e^{itX})E(e^{itY}) = \varphi_X(t)\varphi_Y(t). \qquad \square$$

次の定理は，特性関数と積率の関係を述べたものである．

定理 2–6.2.　ある自然数 n について，$E(|X|^n) < \infty$ であると仮定する．このとき，X の特性関数 $\varphi_X(t)$ は n 回微分可能で，

$$\varphi_X^{(k)}(t) = \frac{d^k}{dt^k}\varphi_X(t) = E\{(iX)^k e^{itX}\}, \quad i = 0, 1, \ldots, n.$$

ただし，$\varphi_X^{(0)}(t) = \varphi_X(t)$ である．特に，

$$E(X^k) = \frac{1}{i^k}\varphi_X^{(k)}(0), \quad k = 0, 1, \ldots, n,$$

が成り立つ．

証明　積分論におけるヘルダーの不等式（[1]参照）を利用すると，

$$E|(iX)^k e^{itX}| \leq E|X|^k \leq (E|X|^n)^{k/n} < \infty$$

となるので，$E\{(iX)^k e^{itX}\}$ は有限確定する．以下，定理の結論を帰納法で示す．$k=0$ については，特性関数の定義そのものなので明らか．いま，結論が $k(\leq n-1)$ まで成り立つと仮定する．

$$\frac{\varphi_X^{(k)}(t+h)-\varphi_X^{(k)}(t)}{h} = E\left\{(iX)^k e^{itX}\frac{e^{ihX}-1}{h}\right\}. \tag{2・1}$$

指数関数のテイラー展開を用いると，不等式

$$\left|\frac{e^{ihx}-1}{h}\right| \leq |x|$$

が成り立つことが分かるので，

$$\left|(iX)^k e^{itX}\frac{e^{ihX}-1}{h}\right| \leq |X|^{k+1}$$

かつ $E|X|^{k+1} \leq (E|X|^n)^{(k+1)/n} < \infty$. したがって，(2・1) にルベーグの収束定理（[1]参照）を適用して，

$$\varphi_X^{(k+1)}(t) = \lim_{h\to 0}\frac{\varphi_X^{(k)}(t+h)-\varphi_X^{(k)}(t)}{h} = E\left\{\lim_{h\to 0}(iX)^k e^{itX}\frac{e^{ihX}-1}{h}\right\} = E\{(iX)^{k+1}e^{itX}\}.$$

よって，$k+1(\leq n)$ でも成り立つことが分かる．定理の後半は，上で示した $\varphi_X^{(k)}(t)$ の期待値による表現式に $t=0$ を代入すれば明らかである． □

特性関数は積率を計算するだけでなく，2 つの確率変数の確率分布が等しいか否かを確かめることにも利用できる．次の結果は，特性関数と確率分布関数の関係を述べたものである．

定理 2–6.3. $F_X(x)$ と $\varphi_X(t)$ をそれぞれ確率変数 X の確率分布関数と特性関数とする．このとき，$F_X(x)$ の任意の連続点 $a,b \in R$ について，

$$F_X(b)-F_X(a) = \lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-ibt}-e^{-iat}}{-it}\varphi_X(t)\,dt$$

が成り立つ．

証明は煩雑なので省略する（[2], [3]参照）．

2 つの確率変数の確率分布が等しければ，確率関数（離散型の場合）または確率密度関数（連続型の場合）も等しい．したがって，特性関数の定義から，この場合はそれぞれの確率変数の特性関数も互いに等しくなる．よって，本章定理 2–6.3 とあわせて，次の結果を得る．

定理 2–6.4. X,Y を確率変数とし，$\varphi_X(t),\varphi_Y(t)$ をそれぞれの特性関数とする．このとき，

$$X \text{ と } Y \text{ の確率分布が等しい} \Leftrightarrow \text{任意の } t \in R \text{ について } \varphi_X(t) = \varphi_Y(t).$$

また，特性関数は，確率分布関数の収束を確かめる方法としても使うことができる．次の定理はその事実を保証するもので，3 編 3 章で述べる中心極限定理の証明にも用いられる．

定理 2–6.5. $X, X_1, X_2, \ldots$ を確率変数，$F_X(x), \varphi_X(t)$ をそれぞれ X の確率分布関数と特性関数，$F_{X_n}(x), \varphi_{X_n}(t)$ をそれぞれ X_n の確率分布関数と特性関数とする．このとき，次が成り立つ．

$$\begin{aligned} & F_X \text{ のすべての連続点 } x \text{ について } \lim_{n\to\infty} F_{X_n}(x) = F_X(x) \\ \Leftrightarrow \quad & \text{任意の } t \in R \text{ について } \lim_{n\to\infty} \varphi_{X_n}(t) = \varphi_X(t). \end{aligned}$$

証明は省略する（[3] 参照）．

特性関数は確率・統計の分野で非常に便利な道具であるが，その計算に複素数の無限級数や複素積分を必要とするところが若干厄介である．そこで，特性関数とほぼ同じ役割を果たすことができる実関数を導入する．

定義 2–6.2. X を確率変数とし，t についての実数値関数 $E(e^{tX})$ について，集合 $D = \{t : E(e^{tX}) < \infty\}$ が 0 を内点に含むとき，

$$m_X(t) = E(e^{tX})$$

を X の**積率母関数** (moment generating function) と呼ぶ

ここで注意すべき点は，0 が D の内点でなければ積率母関数は定義されないということである．特性関数は常に存在するが，積率母関数は必ずしもそうではない．しかし，確率変数 X について積率母関数 $m_X(t)$ が存在するならば，X の特性関数 $\varphi_X(t)$ との間には

$$m_X(t) = \varphi_X(-it), \quad \varphi_X(t) = m_X(it)$$

という関係が成り立つ．見た目には，特性関数と積率母関数の違いは虚数単位の有無だけなので，特性関数で成り立つことは，およそ積率母関数でも成り立つと想像できる．実際，次の定理が成り立つ．

定理 2–6.6. X, Y を確率変数とし，それらの積率母関数 $m_X(t), m_Y(t)$ が存在すると仮定する．このとき，以下のことが成り立つ．

(1) $X+Y$ の積率母関数 $m_{X+Y}(t)$ が存在すると仮定する．このとき，

$$X \text{ と } Y \text{ が互いに独立} \Rightarrow m_{X+Y}(t) = m_X(t)m_Y(t).$$

(2) ある自然数 n について $E|X|^n < \infty$ ならば，$m_X(t)$ は $t=0$ の近傍で n 回微分可能であり，

$$m_X^{(k)}(t) = \frac{t^k}{dt^k} m_X(t) = E(X^k e^{tX}), \quad k = 0, 1, \ldots, n.$$

ただし，$m_X^{(0)}(t)=m_X(t)$ である．特に，

$$E(X^k)=m_X^{(k)}(0),\quad k=0,1,\ldots,n,$$

が成り立つ．

(3) X と Y の確率分布が等しい $\Leftrightarrow$ 任意の $t\in R$ について $m_X(t)=m_Y(t)$.

証明は，特性関数の場合とほとんど同様にしてできる．

■12 群 - 3 編 - 2 章

2-7 条件付き分布

（執筆者：加藤　剛）［2009 年 3 月 受領］

3 編 1 章**??**節において，条件付き確率を定義し，それを用いて事象の独立性を定義した．同様の概念を確率関数や確率密度関数を対象にして考え，確率変数の独立性を定義する．

定義 2–7.1. X,Y を確率変数とする．

(i) X と Y が離散型確率変数で，$X \in D_X = \{x_0, x_1, \ldots\}, Y \in D_Y = \{y_0, y_1, \ldots\}$ とする．また，$p_{X,Y}(x,y)$ を同時確率関数，$p_Y(y)$ を Y の周辺確率関数とする．このとき，

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}, \quad x \in D_X,$$

を，$Y = y \in D_Y$ が与えられたときの X の**条件付き確率関数** (conditional probability function) と呼ぶ．

(ii) X と Y が連続型確率変数で，$f_{X,Y}(x,y)$ を同時確率密度関数，$f_Y(y)$ を Y の周辺確率密度関数とする．このとき，1 つの固定された y に対して，

$$f_{X|Y}(x|y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)} & : \quad f_Y(y) > 0 \\ 0 & : \quad f_Y(y) = 0 \end{cases}, \quad x \in R,$$

を，$Y = y$ が与えられたときの X の**条件付き確率密度関数** (conditional probability density function) と呼ぶ．

条件付き確率関数または条件付き確率密度関数を使うと，**条件付き確率** (conditional probability) が，例えば次のように求められる．

$$P\{X \in A | Y = y\} = \sum_{\{i : x_i \in A\}} p_{X|Y}(x_i|y)$$

: $Y = y$ が与えられたときの離散型確率変数 $X \in A$ の条件付き確率.

$$P\{a < X \le b | Y = y\} = \int_a^b f_{X|Y}(x|y)\,dx$$

: $Y = y$ が与えられたときの連続型確率変数 $a < X \le b$ の条件付き確率.

条件付き確率関数または条件付き確率密度関数を用いると，2 つの確率変数 X と Y の独立性は，次の形で述べられる．

定理 2–7.1.

(1) X,Y が離散型確率変数で，$X \in D_X = \{x_0, x_1, \ldots\}$, $Y \in D_Y = \{y_0, y_1, \ldots\}$ とする．また，$p_{X|Y}(x|y)$ を条件付き確率関数，$p_X(x)$ を X の周辺確率関数とする．このとき，

$$X \text{ と } Y \text{ が互いに独立} \Leftrightarrow \text{任意の } x \in D_X, y \in D_Y \text{ について } p_{X|Y}(x|y) = p_X(x).$$

(2) X,Y が連続型確率変数で，$f_{X|Y}(x|y)$ を条件付き確率密度関数，$f_X(x)$ を X の周辺確率密度関数とする．このとき，

$$X \text{ と } Y \text{ が互いに独立} \Leftrightarrow \text{任意の } x,y \in R \text{ について } f_{X|Y}(x|y) = f_X(x).$$

証明 (2) について示す．(1) も同様である．本章定理 2–4.1 の (3) を利用して，

$$\begin{aligned} X \text{ と } Y \text{ が互いに独立} \quad &\Leftrightarrow \quad \text{任意の } x,y \in R \text{ について } f_{X,Y}(x,y) = f_X(x) f_Y(y) \\ &\Leftrightarrow \quad \text{任意の } x,y \in R \text{ について } f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = f_X(x). \end{aligned}$$ □

本章定義 2–5.1 で見たように，確率関数と確率密度関数が与えられると期待値を定めることができた．条件付き確率関数と条件付き確率密度関数に対しても，同様に条件付き期待値という概念を導入することができる．

定義 2–7.2.

(i) X,Y が離散型確率変数で，$X \in D_X = \{x_0, x_1, \ldots\}$, $Y \in D_Y = \{y_0, y_1, \ldots\}$ とする．また，$p_{X|Y}(x|y)$ を条件付き確率関数とするとき，

$$E(X|Y=y) = \sum_{i=0}^{\infty} x_i \, p_{X|Y}(x_i|y)$$

を，$Y = y \in D_Y$ が与えられたときの X の**条件付き期待値** (conditional expectation) と呼ぶ．

(ii) X,Y が連続型確率変数で，$f_{X|Y}(x|y)$ を条件付き確率密度関数とするとき，

$$E(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx$$

を，$Y = y$ が与えられたときの X の**条件付き期待値** (conditional expectation) と呼ぶ．

条件付き期待値 $E(X|Y=y)$ は y 次第で異なる値をとるので，$E(X|Y=y)$ 自体が y の関数とみなせる．そこで，

$$Y(\omega) = y \Rightarrow E(X|Y)(\omega) = E(X|Y=y)$$

と定めると，$E(X|Y)(\omega)$ は可測空間 $(\Omega, \mathscr{F})$ 上の確率変数となる．

条件付き期待値は，次の性質を持つ．

定理 2–7.2. X,Y,Z を確率変数，a,b を任意の実数とする．

(1)　$E(a|Y) = a.$

(2)　$E(aX+bY|Z) = aE(X|Z)+bE(Y|Z).$

(3)　$E\{E(X|Y)\} = E(X).$

(4)　$g(y)$ を R 値ボレル可測関数とするとき，$E(g(Y)X|Y) = g(Y)E(X|Y).$

証明 X,Y,Z が連続型の場合を証明する．離散型の場合も，確率関数を使えば同様に示せる．
(1) $Y=y$ が任意に与えられたとき，

$$E(a|Y=y) = \int_{-\infty}^{\infty} a\, f_{X|Y}(x|y)\, dx = \frac{a}{f_Y(y)} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx = a\frac{f_Y(y)}{f_Y(y)} = a.$$

$Y=y$ は任意だったので，$E(a|Y) = a.$
(2) $f_{X,Y,Z}(x,y,z)$ を (X,Y,Z) の同時確率密度関数とすると，$(X,Z),(Y,Z),Z$ の周辺確率密度関数，及び，$Z=z$ が与えられたときの (X,Y) の条件付き確率密度関数は，それぞれ

$$\begin{aligned}
f_{X,Z}(x,z) &= \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dy, \\
f_{Y,Z}(y,z) &= \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dx, \\
f_Z(z) &= \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dxdy, \\
f_{(X,Y)|Z}(x,y|z) &= \frac{f_{X,Y,Z}(x,y,z)}{f_Z(z)}
\end{aligned}$$

として得られる．よって，$Z=z$ が任意に与えられたとき，

$$\begin{aligned}
& E(aX+bY|Z=z) \\
= \;& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ax+by)\, f_{(X,Y)|Z}(x,y|z)\, dxdy \\
= \;& a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f_{(X,Y)|Z}(x,y|z)\, dxdy + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{(X,Y)|Z}(x,y|z)\, dxdy \\
= \;& a\int_{-\infty}^{\infty} \frac{x}{f_Z(z)} \left(\int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dy\right) dx + b\int_{-\infty}^{\infty} \frac{y}{f_Z(z)} \left(\int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dx\right) dy \\
= \;& a\int_{-\infty}^{\infty} x\, \frac{f_{X,Z}(x,z)}{f_Z(z)}\, dx + b\int_{-\infty}^{\infty} y\, \frac{f_{Y,Z}(y,z)}{f_Z(z)}\, dy \\
= \;& a\int_{-\infty}^{\infty} x f_{X|Z}(x|z)\, dx + b\int_{-\infty}^{\infty} y f_{Y|Z}(y|z)\, dy \\
= \;& aE(X|Z=z) + bE(Y|Z=z).
\end{aligned}$$

(3) 条件付き確率密度関数の定義に基づいて計算すると，

$$
\begin{aligned}
E\{E(X|Y)\} &= \int_{-\infty}^{\infty} E(X|Y=y)\, f_Y(y)\, dy = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx \right) f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)\, dy \right) dx = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \right) dx \\
&= \int_{-\infty}^{\infty} x f_X(x)\, dx = E(X).
\end{aligned}
$$

(4) $y \in R$ を任意に与えたとき，$g(y)$ は定数となる．そこで，$Y(\omega) = y$ を満たす ω について，(2) を利用すると，

$$
E(g(Y)X|Y)(\omega) = E(g(y)X|Y=y) = g(y)E(X|Y=y) = g(Y(\omega))E(X|Y)(\omega).
$$

y は任意であったから，$E(g(Y)X|Y) = g(Y)E(X|Y)$. □

■参考文献

1) 梅垣壽春, 大矢雅則, 塚田真 :“ 測度・積分・確率, ” 共立出版, 1987.
2) 河田龍夫 :“ フーリエ解析と統計, ” 共立出版, 1985.
3) 西尾真喜子 :“ 確率論, ” 実教出版, 1981.
4) 鈴木　武, 山田作太郎 :“ 数理統計, ” 内田老鶴圃, 1996.